Automated breast masses segmentation in digitized mammograms

Zhang, Han; Foo, Say Wei; Thng, Choon Hua; Krishnan, Shankar M.

2004

Zhang, H., Foo, S. W., Krishnan, S. M., & Thng, C. H. (2004). Automated breast masses segmentation in digitized mammograms. IEEE International Workshop on BioMedical Circuits & Systems (pp. 1-4).

<https://hdl.handle.net/10356/90926>

<https://doi.org/10.1109/BIOCAS.2004.1454102>

© 2004 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder. <http://www.ieee.org/portal/site>.

AUTOMATED BREAST MASSES SEGMENTATION IN DIGITIZED MAMMOGRAMS

Han Zhang¹, Say Wei Foo¹, S.M. Krishnan¹ and Choon Hua Thng²

¹School of Electrical and Electronics Engineering, Nanyang Technological University

²National Cancer Center, Singapore

ABSTRACT

In this paper, an automated segmentation method is proposed. The method is applied to the segmentation of breast masses in digitized mammograms and it operates on the whole mammograms instead of manually selected regions. Pixels with local maximum gray levels are flagged as seeds, from which many candidate objects are grown using modified region-growing technique. Following which False Positive (FP) reduction using decision tree is applied to discard the normal tissue regions. A total of 40 mammograms from Mammographic Image Analysis Society (MIAS) are analyzed. 36 masses are correctly segmented by the proposed method, resulting in 90% True Positive Rate at 1.3 FPs per image.

1. INTRODUCTION

Breast cancer has been the most common cancer among women in Singapore over the past 30 years [1]. Early diagnosis and treatment can increase the chances of survival. Early detection of breast cancer starts with qualitative analysis of mammographic data by physician. Mammography is currently the only proven and cost-effective method to detect breast cancer in its early state of growth. Computer-Aided-Detection (CADetection) has been employed to alert radiologists and the approach serves as a second reader that can be taken into consideration in making the final interpretation.

Mass segmentation is often used as the first step in many CADiagnosis systems for classification of breast masses as malignant or benign. Although some research works have been done on mass segmentation, most of them approach the problem by segmenting the masses in manually separated rectangular areas containing the masses [2-5], or only focusing on one particular kind of mass, e.g. circumscribed lesions [6], and spiculated masses [7]. Segmenting masses from the whole mammographic images is only studied by a few researchers. This category of segmentation has practical meaning for CADiagnosis system for breast cancer [8].

In this paper, an automated mass segmentation method is proposed, which operates on the whole mammograms, and results in a number of regions considered being the most suspicious. The structure of this paper is as follows. The proposed method is described in Section 2. In Section 3, results of

experiments together with discussion of the results are presented. Finally the conclusion is drawn in Section 4.

2. PROPOSED METHOD

2.1. Mass Segmentation Scheme

The flowchart of the proposed method is shown in Fig. 1. Since the mammograms in the MIAS database are clipped/padded to 1024×1024 pixels, there are some irrelevant areas along the margins of the images, e.g. labels, scanning artefacts. Therefore, the first step is to delete these irrelevant regions, focusing on the breast region. Next, the pixels with maximum gray values in a window of 31×31 pixels are flagged as local maxima and each of the local maxima is taken as a seed, from which a candidate object is grown using a modified region-growing technique. After that, a decision tree is employed to reject the objects with normal tissue. The outputs are a number of segmented regions considered being the most suspicious.

Fig. 1. Flowchart of the proposed breast masses segmentation method

The proposed method operates on the whole mammograms, in contrast to most other methods that operate only on manually isolated areas that contain the masses [2-5]. By choosing appropriate parameters in advance, the method is able to automatically isolate the most suspicious mass regions without any human intervention during the procedure.

2.2. Modified Region Growing

Region growing [9] is one of the most popular methods to segment regions in images. Traditional method uses only a constant as the threshold of gray level difference to control the procedure of adding pixels. In respect that mammographic images containing lesions are always characterized with complex background tissue, fuzzy boundary, and low contrast, three criteria were used in this study to control the procedure of adding pixels.

$$|I(x, y) - I(rx, ry)| < t \times I(rx, ry) \quad (1)$$

$$|\sigma_{new} / \mu_{new} - \sigma / \mu| \leq T1 \quad (2)$$

$$|G(x, y)| \leq T2 \quad (3)$$

where $I(x, y)$ is the gray value of the candidate 8-connected neighbors, $I(rx, ry)$ is that of the pixel in the grown region, and $t \times I(rx, ry)$ provides an adaptive threshold of gray level difference, which is a big difference from traditional region growing method. In the experiment t is selected from 0.01 to 0.03. σ , μ are the gray level standard deviation and mean of the grown region, while σ_{new} , μ_{new} are those of the new region when the candidate pixel is added. The threshold $T1$ indicates the desired degree of homogeneity between the old subregion and the future subregion. $G(x, y)$ is the gradient magnitude of candidate in the 8-connected pixels. The gradient threshold $T2$ guarantees that the growing procedure will stop when it reaches the pixels with high gradient value which can be regarded as the mass margins.

2.3. False Positive Reduction

Since the mammograms containing lesions are usually characterized by low-contrast, fuzzy boundary of the lesions as well as complex background, a lot of normal tissue objects are included as suspected lesions after the initial region growing. False Positive (FP) reduction has to be carried out in order to discard these normal tissue regions. In this paper, a decision tree approach is employed to reject the objects with normal tissue. Seven geometric features are used to discriminate normal tissue from masses. These are: radius, circularity, compactness, contrast, tumour boundary roughness (TBR), mean and standard deviation of each segmented region.

$$radius = \frac{\sum r_d}{M} \quad (4)$$

$$circularity = \frac{\mu_R}{\sigma_R} \quad (5)$$

$$compactness = \frac{P^2}{A} \quad (6)$$

$$contrast = \frac{g_{in}}{g_{out}} \quad (7)$$

$$TBR = \frac{N_m}{N_t} \quad (8)$$

$$mean = \frac{1}{N} \sum_{(i,j) \in R} g(i, j) \quad (9)$$

$$std = \frac{1}{N} \sum_{(i,j) \in R} (g(i, j) - g_{avg})^2 \quad (10)$$

where r_d is the radial distance from centroid to each boundary point, and M is the total number of the point that make up of the region boundary. μ_R and σ_R is the mean and the standard deviation of radial distance of the region boundary respectively. P is the length of the object's perimeter, and A is the object's area. g_{in} is the mean gray level of the segmented region, and g_{out} is the mean gray level of the 2-pixel width boundary. N_m is the number of angles in the boundary perimeter which have multiple boundary points at a given angle, N_t is the total number of angles that make up the boundary. $g(i, j)$ is the gray level of pixel (i, j) inside the segmented object region, and N is the number of pixels inside the segmented region.

3. RESULTS AND DISCUSSION

3.1. Data Set

For our experiments, the MIAS MiniMammographic Database [10] provided by Mammographic Image Analysis Society (MIAS) was used. The mammograms in this database are digitized at 200-micron pixel edge and clipped/padded so that every image is 1024 × 1024 pixels. The MIAS database provides ground-truth for each mass in the form of approximate radius of a circle enclosing the mass.

Forty mammographic images containing biopsy-proven masses are selected to evaluate the algorithm. The approximate radius of a circle enclosing the mass in our data set ranges from 17 pixels to 174 pixels. Table I shows the statistics of data set used in this study.

Table I. Statistics of Data Set

Background tissue Severity	Fatty	Fatty-Glandular	Dense	Total
Benign	11	5	5	21
Malignant	10	7	2	19
Total	21	12	7	40

3.2. Segmentation Results

For the validation of the mass segmentation method, the criterion of overlap is used.

$$overlap = \frac{area\{S \cap U\}}{area\{S \cup U\}} \quad (11)$$

where S is the region of the segmented object, U is that of the ground-truth mass.

Since the ground-truth provided by the database is the approximate radius of a circle enclosing the mass, not the accurate mass boundary drawn by experts, the overlap criterion is set to a small value of 0.30. Segmented region with overlap greater than 0.30 is regarded as a True Positive (TP), otherwise a FP. The proposed method detected correctly 36 out of 40 mammograms with masses that satisfied the above overlap criterion, resulting in 90% True Positive Rate.

The method failed to detect the masses in the remaining 4 mammograms, and 3 of them were not grown into objects with appropriate size because of discontinuity of the masses, thus being rejected in the step of FP reduction. In the case of Mammogram mdb141, shown in Fig. 2, the seed inside the mass was not found during the process of detection of local maxima.

In the 36 successfully segmented mammograms, there are 1.3 False Positives per image. However, by setting less stringent FP reduction criteria, the TP rate can be increased at the expense of higher FPs per image.

Fig. 3 shows an example of mass segmentation using this algorithm. Mammogram mdb010 contains a benign circumscribed mass embedded in fatty background tissue. The true mass is successfully segmented without any False Positive.

Mammogram mdb091 shown in Fig. 4 also contains a benign circumscribed mass embedded in fatty background tissue. However, this mass cannot be easily recognized by human eyes compared with mdb010, because the mass has a fuzzy boundary. However, with the proposed method, the true mass is successfully detected without any False Positive. The segmentation result is shown in Fig. 4(b).

Mammogram mdb145 contains a benign spiculated mass embedded in dense background tissue. Its segmentation result is shown in Fig. 5(b). One normal tissue object, i.e. a FP, is retained even after FP reduction.

Fig. 2. mdb141

(a)

(b)

(c)

(d)

Fig. 3. mdb010 (a).original digitized mammogram; (b). Breast region of mdb010; (c). Local maxima (marked with dark points); (d). Segmentation result after False Positive reduction.

(a)

(b)

Fig. 4. mdb091 (a). Local maxima; (b). Segmentation result

(a)

(b)

Fig. 5. mdb145 (a). Local maxima; (b). Segmentation result.

4. CONCLUSION

In this paper an automated mass segmentation method is proposed. The method is able to deal with mammograms as a whole instead of just dealing with isolated areas containing masses. By choosing appropriate parameters in advance, the method is able to automatically isolate the most suspicious mass

regions without any human intervention in the loop. The method is applied to analyze 40 digitized mammograms containing biopsy-proven masses and a True Positive rate of 90% at 1.3 FPs per image is achieved.

The features computed in this work can be further improved and the region-growing algorithm can be optimized by the addition of conditions to control the procedure.

5. REFERENCES

- [1] ST Quek, CH Thng, JBK Khoo and WL Koh, "Radiologists' detection of mammographic abnormalities with and without a computer-aided detection system," *Australasian Radiology*, 2003, 47, pp. 257-260.
- [2] G. M. te Brake and N. Karssemeijer, "Segmentation of suspicious densities in digital mammograms," *Med. Phys.*, vol. 28, pp. 259-266, 2001.
- [3] H. Kobatake, M. Murakami, H. Takeo, and S. Nawano, "Computerized detection of malignant tumors on digital mammograms," *IEEE Trans. Med. Imag.*, vol. 18, pp. 369-378, 1999.
- [4] M.A. Kupinski, M.L. Giger, "Automated seeded lesion segmentation on digital mammograms," *IEEE Trans. Med. Imag.*, vol. 17, pp. 510-517, 1998.
- [5] D. Guliato, R. M. Rangayyan, W. A. Carnielli, J. A. Zuffo, and J. E. L. Desautels, "Segmentation of breast tumors in mammograms by fuzzy region growing," in *Proc. 20th Annu. Int. Conf. IEEE Eng. in Medicine and Biology Society*, pp. 1002-1005, 1998.
- [6] I. Christoyianni, E. Dermatas, and G. Kokkinakis, "Automatic detection of abnormal tissue in mammography," in *Proc. Int. Conf. on Image Processing*, vol. 2, pp. 877-880, Oct 2001.
- [7] N. Karssemeijer, G.M. te Brake, "Detection of stellate distortions in mammograms," *IEEE Trans. Med. Imag.*, vol. 15, pp. 611 -619, Oct. 1996.
- [8] B. Sahiner, N. Petrick, H. P. Chan, L.M. Hadjiiski, C. Paramagul, M.A. Helvie, M.N. Gurcan, "Computer-aided characterization of mammographic masses: accuracy of mass segmentation and its effects on characterization," *IEEE Trans. Med. Imag.*, vol. 20, pp. 1275 -1284, Dec. 2001.
- [9] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd edition, Prentice-Hall, NJ, USA, 2002.
- [10] <http://www.wiau.man.ac.uk/services/MIAS/MIASmini.html>